

Ambiguity Drives Higher-Order Pavlovian Learning

Running Head: Ambiguity Drives Higher-Order Learning

Tomislav D. Zbozinek, Ph.D.^a

Omar Perez, Ph.D.^a

Toby Wise, Ph.D.^a

Michael Fanselow, Ph.D.^b

Dean Mobbs, Ph.D.^a

- a. California Institute of Technology, Humanities and Social Sciences, 1200 E. California Blvd., MC 228-77, Pasadena, CA 91125, USA
- b. University of California, Los Angeles, Department of Psychology, 1285 Franz Hall, Los Angeles, CA 90025

Correspondence to:

Tomislav D. Zbozinek, Ph.D.

Division of Humanities and Social Sciences

California Institute of Technology

1200 E. California Blvd., MC 228-77

Pasadena, CA 91125, USA

Email: zbozinek@caltech.edu

Abstract

In the natural world, stimulus-outcome associations are often noisy and ambiguous. Learning to disambiguate these associations to identify which specific outcomes will occur is critical for survival. Pavlovian occasion setters are stimuli that determine whether other stimuli that are ambiguous will result in a specific outcome. Occasion setting is a well-established field, but very little investigation has been conducted on how occasion setters are disambiguated when they themselves are ambiguous. We investigated the role of higher-order Pavlovian occasion setting in humans. We also developed and tested the first computational model predicting direct associations, traditional occasion setting, and 2nd-order occasion setting. Results showed that occasion setters affected ambiguous but not unambiguous lower-order stimuli and that 2nd-order occasion setting was indeed learned. Our computational model demonstrated excellent fit with the data, advancing our theoretical understanding of learning with ambiguity. These results may ultimately improve treatment of Pavlovian-based mental health disorders (e.g., anxiety).

Survival requires the organism to quickly learn which situations predict safety vs danger and reward vs no reward. This is straightforward to accomplish when a given situation consistently predicts a specific outcome (e.g., ¹). However, when the situation predicts more than one outcome (and is therefore ambiguous), it can be difficult to determine which outcome will occur, which may have survival consequences. The most prominent experimental paradigm for learning stimulus-outcome associations is Pavlovian conditioning, in which the organism learns associations between conditional stimuli (CSs) and their outcomes (i.e., unconditional stimuli; USs). The CS+ is a CS that is paired with the US, whereas the CS- is paired with the absence of the US. These tend to be *unambiguous* stimuli – meaning, they always (or almost always) predict their respective outcomes. However, in the real world, it is rare for a stimulus to be truly unambiguous. Rather, stimuli tend to be *ambiguous* since they sometimes predict one outcome and other times predict other outcomes depending on the situation. In order to optimize survival, the organism must learn to disambiguate these stimuli/situations and predict their outcomes accurately.

For example, perhaps a specific child usually behaves well, but the parent is interested in figuring out which situations lead to bad behavior from the child. Because the child often behaves well but sometimes behaves badly, there is an ambiguous association between the child and bad behavior. From a Pavlovian perspective, the parent needs to learn which stimuli, situational factors, or contexts determine whether the child (CS) will behave badly (US). One of the primary Pavlovian experimental designs investigating ambiguous CSs is occasion setting (²⁻⁹), in which an occasion setter is a stimulus that indicates whether the CS will result in the US. In this case, perhaps the child has a friend that is a bad influence, and when the child sees this friend in the afternoon, the child behaves badly later that night. Theoretically, the friend (occasion setter) would enable the child/bad behavior (CS/US) association, but the absence of the friend would enable the child/good behavior (CS/No US) association. Indeed, occasion setters can either enable or disable the CS/US association. Those that enable it are called positive occasion setters (i.e., the CS predicts the US *only if* the positive occasion setter was presented), and those that disable it are negative occasion setters (i.e., the CS predicts the US *unless* the negative occasion setter was presented).

Occasion setting is thought to operate via modulation (^{2,9}), where the occasion setter directly affects the CS/US association (other theories, such as configuration, show considerable accuracy, though less than modulation; ⁹). The modulation account posits that stimuli are arranged hierarchically in meaning, where higher-order learning (i.e., occasion setting) affects lower-order learning (i.e., direct associations: learning that a CS directly predicts the presence or absence of the US). Multiple tests can be performed to discern whether occasion setting was indeed learned, as opposed to other explanations (e.g., that the “occasion setter” simply formed a direct association with the US; ¹⁰). First, a stimulus can simultaneously possess both direct and occasion setting values (^{2,3,11,12}) – meaning, it can be a CS- (directly inhibiting the US memory) and a positive occasion setter (enabling the CS/US association of a different CS). Using the example above, the friend could directly inhibit bad behavior for themselves but enable bad behavior from the other child. Second, an occasion setter will only influence CSs that have undergone training with an occasion setter (¹¹⁻¹⁷). This means that occasion setters will only affect responding to *ambiguous* CSs, though some specificity is involved (e.g., ^{11-13,16}). These two properties can be tested with a combination of training and transfer test procedures, as are done in the present experiments.

Many studies have been conducted on traditional occasion setting described above (^{2,3,9}; hereafter referred to as 1st-order occasion setting), but there is very limited research on 2nd-order occasion setting, including studies investigating it directly (¹⁸) or indirectly (^{13,19-21}). We define second-order occasion setters as stimuli that determine how *ambiguous* 1st-order occasion setters will affect the CS/US association. Using our example from above, the child (CS) ordinarily does not behave badly (US) unless they see their friend (1st-order occasion setter). Converting this to a 2nd-order occasion setting example, perhaps the friend only *sometimes* makes the child behave badly, making the friend an ambiguous 1st-order occasion setter. A 2nd-order occasion setter would determine whether the friend (1st-order occasion setter) will cause the child

(CS) to behave badly (US). Perhaps the child's grandparent (2nd-order occasion setter) gives the child good advice regarding the friend that protects the child from the friend's bad influence. Thus, if the child (CS) sees the grandparent in the morning, the child will not behave badly (US) that night – even if they see their friend (1st-order occasion setter) in the afternoon. A major goal of the present manuscript is to determine whether humans use 2nd-order occasion setting to resolve ambiguity in stimulus-outcome relationships. We are aware of only one study (using rats) that explicitly attempted to investigate whether 2nd-order occasion setting can be learned (¹⁸). While they successfully trained discriminations that resemble 2nd-order occasion setting, they did not conduct the transfer tests that are necessary to conclude that the underlying learning was indeed 2nd-order occasion setting.

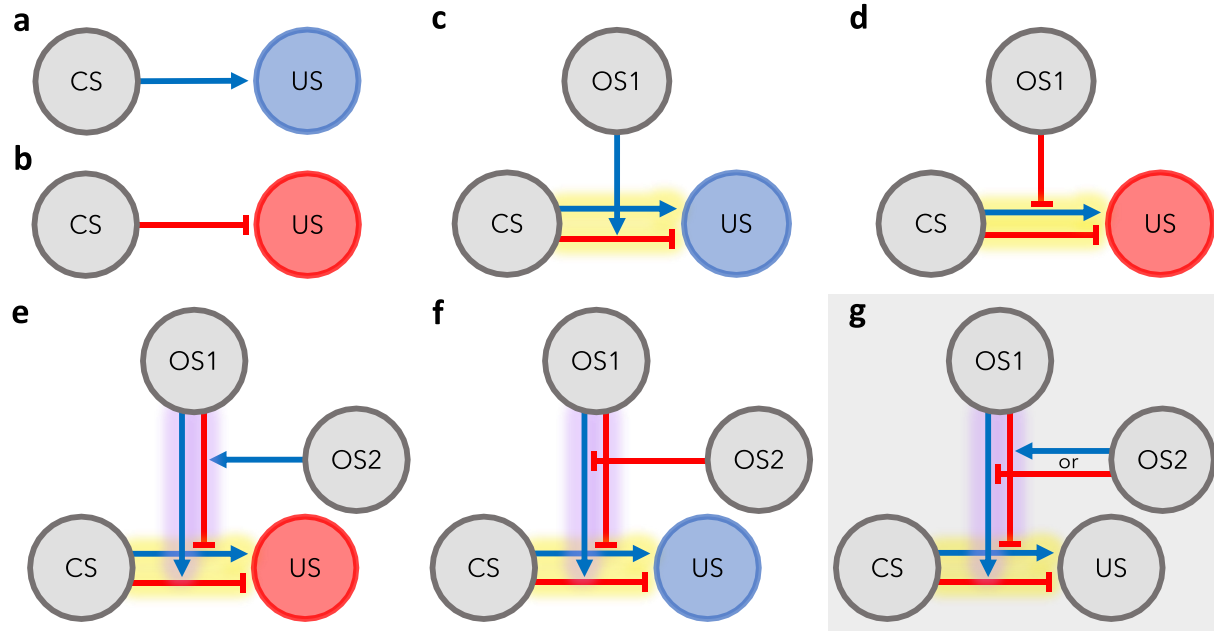


Figure 1. Hierarchical Model of Direct Associative Learning and Occasion Setting. Panels **a-g** display: **a**) direct excitation, **b**) direct inhibition, **c**) 1st-order positive occasion setting, **d**) 1st-order negative occasion setting, **e**) 2nd-order positive occasion setting, **f**) 2nd-order negative occasion setting, and **g**) the total model with all associations. Circles are stimuli: unconditional stimulus (US), conditional stimulus (CS), 1st-order occasion setter (OS1), and second-order occasion setter (OS2). Blue arrows (→) indicate excitatory effects; red line segments (→) indicate inhibitory effects; yellow glow indicates CS ambiguity; purple glow indicates OS1 ambiguity; blue USs indicate US delivery; and red USs indicate US omission. Stimulus ambiguity is the gateway to higher-order associative learning: 1st-order occasion setting is learned only if the CS is ambiguous; 2nd-order occasion setting is only learned if the CS and OS1 are ambiguous. CSs have a direct predictive relationship with the US. If the CS is unambiguous (i.e., consistently predicts the presence *or* absence of the US), no occasion setting is learned. If the CS is ambiguous (i.e., sometimes predicts the US, sometimes predicts absence of the US), then attention is broadened to other stimuli or contextual factors (i.e., OS1) that disambiguate the current meaning of the CS based on the degree to which the CS is ambiguous. OS1 modulates the CS/US association. If OS1 consistently excites the CS/US association, then OS1 is a positive OS1; if OS1 consistently inhibits the CS/US association, then OS1 is a negative OS1. In these cases, OS1 is unambiguous. However, if OS1 sometimes excites and sometimes inhibits the CS/US association (i.e., OS1 is ambiguous), then attention is broadened to other stimuli or contextual factors (i.e., OS2) that disambiguate the current meaning of OS1 based on the degree to which OS1 is ambiguous. Thus, 2nd-order occasion setting is learned only if the CS and OS1 are ambiguous. If OS2 consistently disables OS1's 1st-order positive occasion setting ability, then OS2 is a 2nd-order negative occasion setter. If OS2 consistently disables OS1's 1st-order negative occasion setting ability, then OS2 is a 2nd-order positive occasion setter. Additionally, each hierarchical level (direct associations, 1st-order occasion setting, 2nd-order occasion setting) and excitatory/inhibitory directions are orthogonal – meaning, a given stimulus can be any combination of an excitatory/inhibitory CS, OS1, or OS2 (e.g., a given stimulus can simultaneously be a CS+, negative OS1, and positive OS2).

Thus, while 2nd-order occasion setting is a theoretically plausible learning process, there are no experiments demonstrating it, as the only study attempting to do so ⁽¹⁸⁾ did not administer the appropriate tests to determine whether 2nd-order occasion setting was indeed learned. Additionally, there have been no formal models that predict 2nd-order occasion setting – perhaps because 2nd-order occasion setting has not been explicitly demonstrated. In the present manuscript, we have formalized and tested such a model (see Methods and Supplementary Materials) based upon the idea of hierarchical modulation of associative representations (Figure 1). In our model, lower-level stimulus ambiguity causes learning to hierarchically ascend from a) direct associations to 1st-order occasion setting, and b) 1st-order occasion setting to 2nd-order occasion setting. In other words, if a lower-level stimulus is ambiguous, then higher-order learning can occur.

There were three goals in the present experiments: 1) determine whether 2nd-order occasion setting can be learned, 2) determine whether learning higher-order Pavlovian associations is dependent on lower-order stimulus ambiguity, and 3) evaluate our computational model of 2nd-order occasion setting. To this end, we conducted two mirror-image experiments: a 2nd-order negative occasion setting experiment (Experiment 1) and a 2nd-order positive occasion setting experiment (Experiment 2). To address the first goal, we trained multiple stimuli across discriminations intended to cause learning across three hierarchical levels: direct learning, 1st-order occasion setting, and 2nd-order occasion setting. We conducted specific tests (i.e., transfer tests) to test whether 1st-order and 2nd-order occasion setting were indeed learned. We predicted that each hierarchical level would be orthogonal – meaning, a stimulus could simultaneously signal outcomes in each of the three levels (e.g., Stimulus A: CS+, 1st-order positive occasion setter, 2nd-order negative occasion setter). We also tested whether a 2nd-order occasion setter could affect a CS without the presence of a 1st-order occasion setter, predicting that it could not (i.e., a 2nd-order occasion setter cannot bypass a 1st-order occasion setter to affect a CS directly). To address the second goal, we predicted all occasion setters would *not* affect responding to lower-order stimuli if they were *unambiguous*. Conversely, we predicted 2nd-order occasion setters *would* affect responding to lower-order stimuli if they were *ambiguous* (e.g., ^{13,16,22}). For a list of all specific hypotheses and analyses details, please see our pre-registrations (Experiment 1: <https://osf.io/n2c6v>, Experiment 2: <https://osf.io/hxcfs>). To evaluate our third goal, we conducted computational modeling to evaluate model fit and accuracy using our novel 2nd-order occasion setting model and contrasted its predictions with our 1st-order occasion setting model and the Rescorla-Wagner model ⁽¹⁰⁾.

Results

For full statistical details of results reported below, see Supplemental Materials.

Training

Results from training phases are shown in Figure 3. The critical test of reinforcement learning was the Reminder phases, as this was the end of each training section. Overall, in both experiments, participants correctly learned which stimuli were reinforced and which were not for all trial types – direct associations, 1st-order occasion setting, and 2nd-order occasion setting. The most important and novel of these was 2nd-order occasion setting: as hypothesized, 2nd-order occasion setting trial types had significantly lower (Experiment 1) and greater (Experiment 2) responding than their respective 1st-order occasion setting trial types during Reminder (e.g., Experiment 1: ABC- vs BC+; Experiment 2: UMN+ vs MN-; p s < .001).

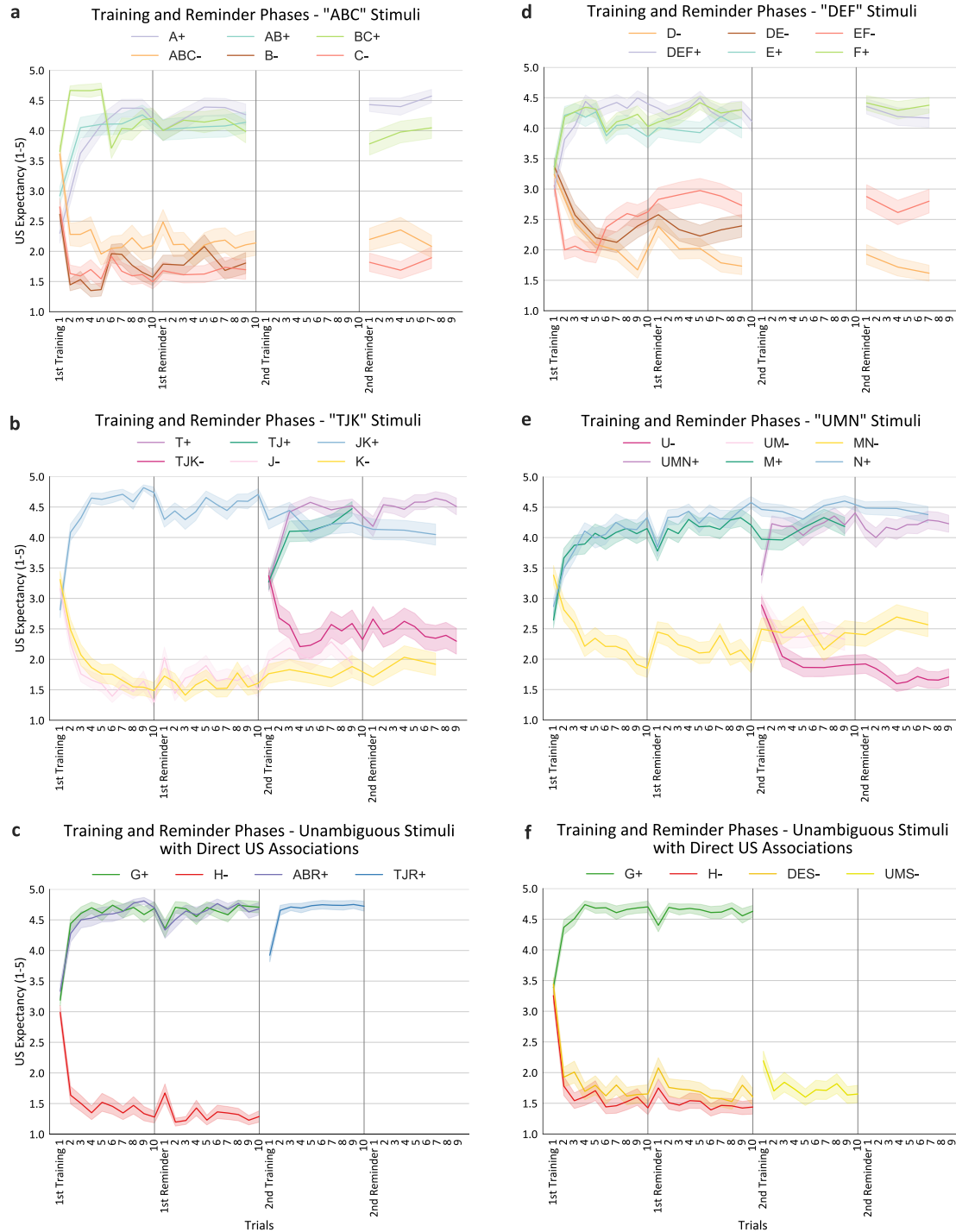


Figure 3. Experiment 1 and 2 Training Results. **a, b, c)** Experiment 1 Training results generally reflect direct CS/US associations, 1st-order positive occasion setting, and 2nd-order negative occasion setting. **d, e, f)** Experiment 2 Training results generally reflect direct CS/US associations, 1st-order negative occasion setting, and 2nd-order positive occasion setting. Congruent conditions/panels are displayed horizontally between experiments. Results in both experiments showed that participants correctly learned which stimuli were (non)reinforced. Error bands reflect standard error. Generally, “cool” colors (blues, greens, purples) indicate hypothesized higher values, whereas “warm” colors (reds, oranges, yellows) indicate hypothesized lower values. See Supplementary Materials for interactive graphs (i.e., html files), where the reader can hide/show conditions, see specific values of data, and zoom in/out.

Transfer Test

See Figure 4 and Supplementary Materials for details on statistical analyses. Overall, all our hypotheses were supported in Experiment 1, and most hypotheses were supported in Experiment 2.

Replicating previous research, we hypothesized that a 1st-order occasion setter would not affect an unambiguous CS. This was tested three times in each experiment. In Experiment 1, all three 1st-order positive occasion setting transfer tests were successful; in Experiment 2, one of three 1st-order negative occasion setting tests was successful (constituting our only two null results across all hypotheses in both experiments). First, in Experiment 1, all three 1st-order occasion setting tests were successful, as evidenced by more similar responding between BH/H- than BH/BC+ (Figure 4a; $p < .001$), JH/H- than JH/JK+ (Figure 4b; $p < .001$), and AH/H- than AH/G+ (Figure 4c; $p < .001$). Congruently, Experiment 2 showed that one of three 1st-order negative occasion setting transfer tests was successful. Specifically, responding to EG was equidistant from EF and G+ (Figure 4f; $p = .588$), and responding to MG was equidistant from MN- and G+ (Figure 4g; $p = .197$). Conversely, our third test showed more similar responding between DG/G+ than DG/H- (Figure 4h; $p < .001$), supporting the hypothesis. See Discussion regarding null results.

Second, as part of our novel hypotheses, we hypothesized that 2nd-order occasion setters could only affect a CS if a 1st-order occasion setter were present; thus, we tested 2nd-order occasion setters on CSs in absence of 1st-order occasion setters, predicting the 2nd-order occasion setter would not affect the CS. Results supported this hypothesis in both experiments. In Experiment 1, AG showed more similar responding to G+ than H- (Figure 4c; $p < .001$), and in Experiment 2, DH showed more similar responding to H- than G+ (Figure 4h; $p < .001$).

Third, one of our critical novel tests was whether a 2nd-order occasion setter would affect *unambiguous* lower-order stimuli; we hypothesized it would not. This was assessed using a) the trained 2nd-order occasion setter/1st-order occasion setter combination with an unambiguous CS, as well as testing b) the 2nd-order occasion setter with a trained unambiguous 1st-order occasion setter/CS combination. In each case and in both experiments,

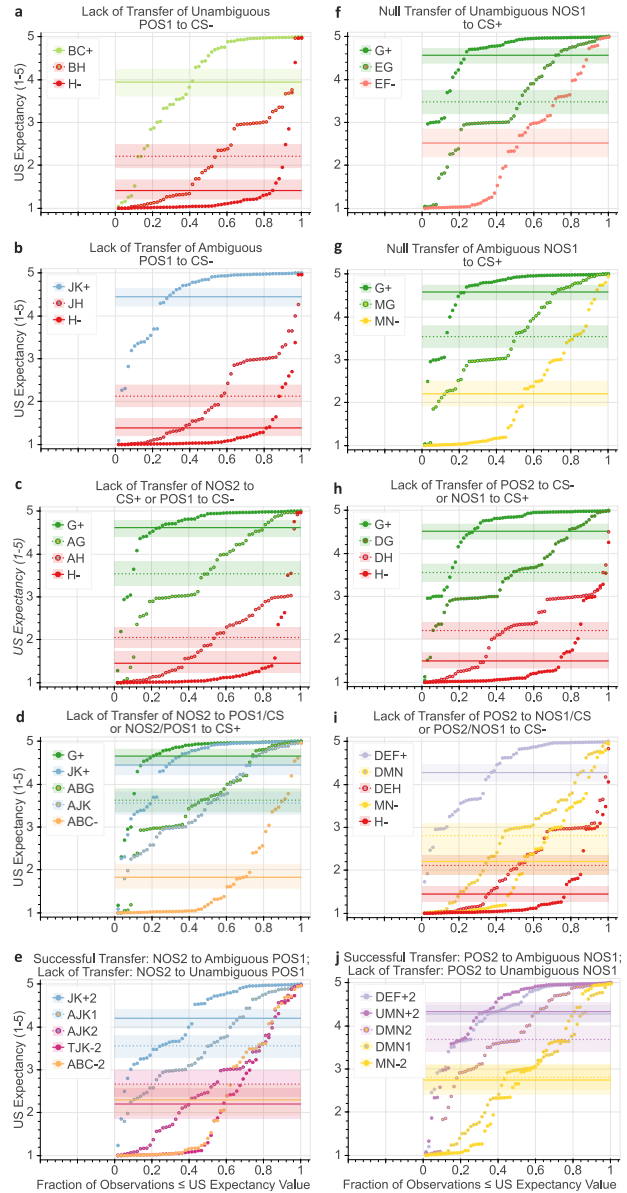


Figure 4. Experiment 1 and 2 Transfer Test Results. Panels a-f are from Experiment 1 (left column); panels g-j are from Experiment 2 (right column). Individual data points are plotted on using an empirical cumulative distribution function, where the X-axis indicates the fraction of observations \leq the US expectancy value. The horizontal lines indicate the mean values of the trained (solid lines) and transfer (dotted lines) stimuli; shaded areas are 95% confidence intervals. "Cool" colors (blues, greens, purples) are hypothesized to have high values, whereas "warm" colors (reds, oranges, yellows) are hypothesized to have low values. Stimuli with mixed colors are transfer stimuli. CS+ = excitatory conditional stimulus; CS- = inhibitory CS; POS1 = 1st-order positive occasion setter; NOS1 = 1st-order negative occasion setter; POS2 = 2nd-order POS; NOS2 = 2nd-order NOS. Most relevant comparisons in panels e (AJK2 vs AJK1) and j (DMN2 vs DMN1) show transfer only occurred with ambiguous stimuli. See Supplementary Materials for interactive graphs (i.e., html files), where the reader can hide/show conditions, see specific values of data, and zoom in/out.

all hypotheses were supported. Specifically, in Experiment 1, AJK had more similar responding to JK+ than ABC- (Figure 4d; $p = .005$), and ABG had more similar responding to G+ than ABC- ($p = .005$). Congruently, in Experiment 2, DMN had more similar responding to MN- than DEF+ (Figure 4i; $p = .002$), and DEH had more similar responding to H- than DEF+ ($p < .001$).

Fourth, our other critical novel test was to evaluate whether the ability of 2nd-order occasion setters to affect lower-order stimuli depended on whether the lower-order stimuli were unambiguous or ambiguous. In the previous paragraph, we demonstrated that 2nd-order occasion setters had little effect on *unambiguous* lower-order OS1/CS combinations. In each experiment, we later trained those same *unambiguous* 1st-order occasion setter/CS combinations with a 2nd-order occasion setter to make them *ambiguous*. We then tested whether a different 2nd-order occasion setter could affect the now-ambiguous 1st-order occasion setter/CS combination more than it did before. Thus, the exact same stimulus combinations (i.e., Experiment 1: AJK; Experiment 2: DMN) were each tested twice – once when the 1st-order occasion setter/CS combinations were unambiguous and later when ambiguous. We hypothesized there would be a greater effect of 2nd-order occasion setters on the 1st-order occasion setter/CS combinations if the latter were *ambiguous*. This hypothesis was supported in both experiments. In Experiment 1, responding was lower to *ambiguous* AJK (i.e., AJK2) than *unambiguous* AJK (i.e., AJK1; Figure 4e; $p < .001$). Congruently, in Experiment 2, responding was higher to *ambiguous* DMN (i.e., DMN2) than *unambiguous* DMN (i.e., DMN1; Figure 4j; $p < .001$).

Computational modeling

While the model-free results above demonstrate the presence of 2nd-order occasion setting, they do not allow us to evaluate the underlying theoretical process through which occasion setters affect expectations (e.g., via lower-order stimulus ambiguity) nor quantitatively evaluate competing theoretical models. To answer this question, we tested a computational model that allowed occasion setters to impact outcome expectations only if lower-order stimuli were ambiguous. That is, the influence of 1st-order occasion setters was dependent on CS ambiguity, and the influence of 2nd-order occasion setters was dependent on 1st-order occasion setter ambiguity and CS ambiguity.

Parameter Recovery

For each model, we simulated random data and evaluated our 2nd- and 1st-order occasion setting models' and the Rescorla-Wagner model's ability to estimate those parameters accurately. In short, all models in both Experiments showed high correlations between simulated and recovered parameters ($r_s > .998$; see Supplemental Materials), indicating that we were able to accurately estimate individual subjects' parameter values.

Model Fit

Results show that, in both experiments, all versions of our 1st- and 2nd-order occasion setting models – which allowed occasion setters to influence expectations of ambiguous CSs – accounted for a significant portion of the variance. They also outperformed the Rescorla-Wagner model (see Figure 5), which makes predictions based purely on direct associations without considering the modulatory effects of occasion setters. Additionally, our 2nd-order occasion setting models outperformed our congruent 1st-order occasion setting models (i.e., when controlling for number of α parameters). Thus, our 2nd-order occasion setting model was able to significantly predict direct learning, 1st-order occasion setting, and 2nd-order occasion setting.

Integrating results from both experiments, the Rescorla-Wagner model accounted for approximately 0% of the variance in the data. Conversely, our 1st-order occasion setting models account for approximately 38% (General α) and 51-54% (Stimulus-Specific α s) of the data. Our best-fitting models were the 2nd-order occasion setting models; among those, the best-fitting models accounted for approximately 60-61% of variance. Additionally, WAIC results were congruent: all occasion setting models outperformed the Rescorla-Wagner model, and the 2nd-order occasion setting models outperformed the 1st-order occasion setting models (controlling for number of α parameters). Overall, collapsing across both experiments, our Stimulus-Specific learning rate model performed the best, which provided a learning rate parameter for each stimulus combination (Experiment 1: WAIC = 29,431, $R^2 = .609$, Experiment 2: WAIC = 30,994, $R^2 = .599$).

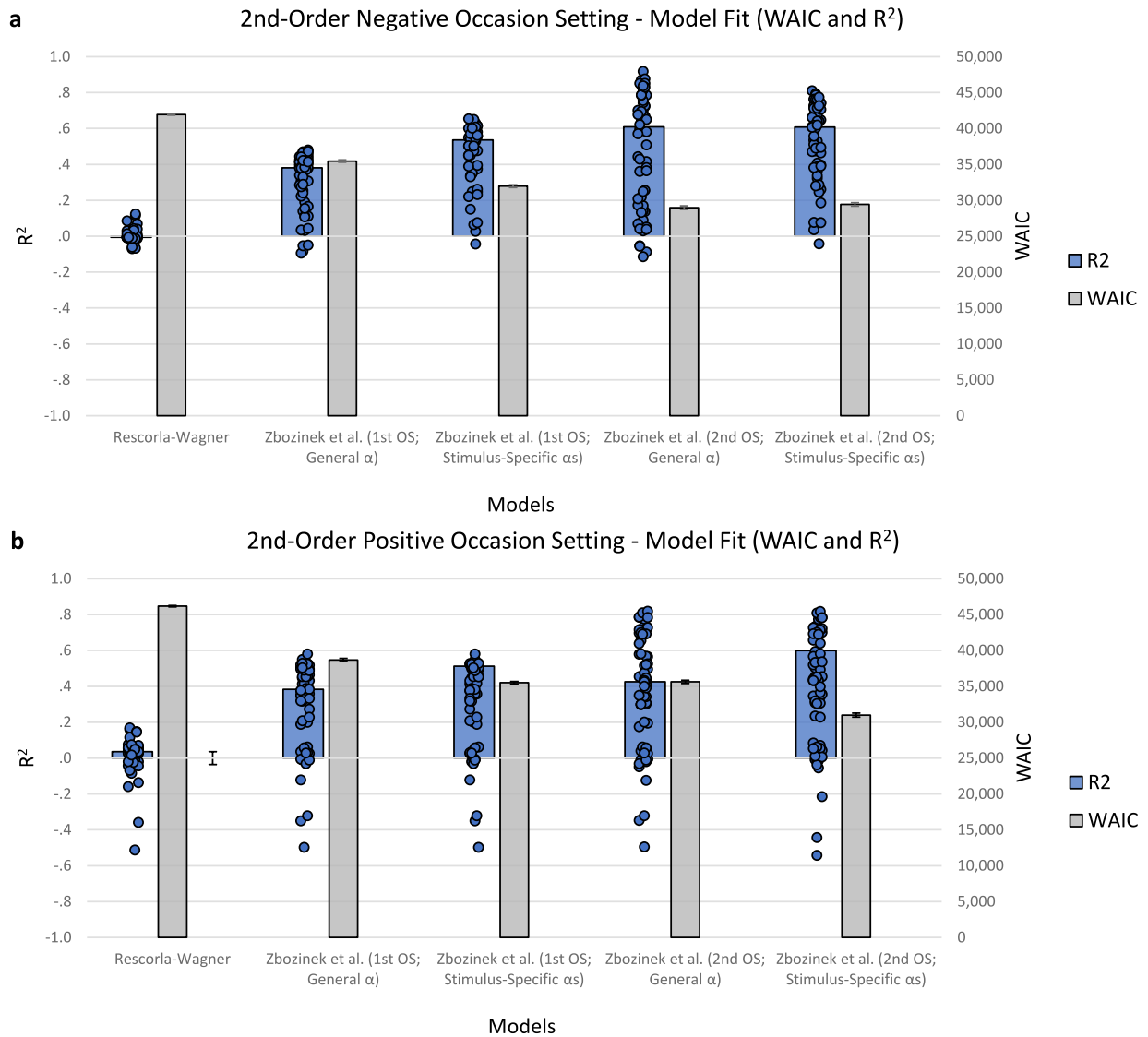


Figure 5. Computational Modeling - Model Fit. Model fit and accuracy were determined with WAIC (lower = better) across all trials and R^2 during Training/Reminder phases. Three models were compared: Rescorla-Wagner model (direct associations only), Zbozinek et al. 1st OS (direct associations and 1st-order occasion setting), and Zbozinek et al. 2nd OS (direct associations, 1st-, and 2nd-order occasion setting). Within the occasion setting models, we tested one general α vs stimulus-specific α s. Bar graph for WAIC indicates mean and standard error. Bar graphs for R^2 values indicate median (due to a few extreme outliers). Individual data points for R^2 values are plotted. R^2 values above 0 indicate variance explained by model; R^2 values below 0 indicate model performed worse than a constant model. Panel **a** is Experiment 1 (2nd-order negative occasion setting); panel **b** is Experiment 2 (2nd-order positive occasion setting). Results show that all occasion setting models outperform the Rescorla-Wagner model in both experiments. Additionally, in Experiment 1 (panel **a**), both 2nd-order occasion setting models outperform the 1st-order occasion setting models (WAIC and R^2) with highly similar results between both 2nd-order occasion setting models. In Experiment 2 (panel **b**), matching for α models, both 2nd-order occasion setting models outperform the 1st-order occasion setting models (WAIC and R^2), and the 2nd-order occasion setting model with stimulus-specific α s outperforms the general α model. Integrating both experiments, the 2nd-order occasion setting model with stimulus-specific α s performs the best (Experiment 1: WAIC = 29,431, R^2 = .609; Experiment 2: WAIC = 30,994, R^2 = .599).

Discussion

This report investigated whether lower-order stimulus ambiguity can cause higher-order Pavlovian learning to occur. This was assessed using an established Pavlovian paradigm that studies stimulus ambiguity: Pavlovian occasion setting (^{2,3,9}). From a hierarchical modulation perspective, traditional 1st-order occasion setting occurs when one stimulus (i.e., the 1st-order occasion setter) determines whether another stimulus (i.e., the conditional stimulus; CS) will result in a specific outcome (i.e., the unconditional stimulus; US). First-order occasion setting is a well-established area of research (^{2,3,9}), but research on the existence of 2nd-order occasion setting is inconclusive and almost non-existent (¹⁸). We hypothesized that a stimulus (i.e., the 2nd-order occasion setter) determines how a 1st-order occasion setter affects the CS/US association only if the lower-order stimuli (i.e., CSs, 1st-order occasion setters) are ambiguous (i.e., when they only sometimes predict the US), not when they are unambiguous (i.e., when they always predict either the presence or absence of the US). This is highly relevant for real-life learning and decision-making, as associations between stimuli/contexts and their outcomes tend to be ambiguous (e.g., whether a public speech will be well-received usually depends on many factors). Our experiments are the first to investigate whether 2nd-order occasion setting exists by conducting tests of the underlying learning processes. The results of our experiments showed strong support for our hypotheses, including that participants used 2nd-order occasion setting to learn about the reinforcement of ambiguous (but not unambiguous) lower-order stimuli (i.e., 1st-order occasion setters, CSs).

First, we assessed whether a 1st-order occasion setter could affect an unambiguous CS. In three tests in Experiment 1, we replicated previous studies showing that a 1st-order positive occasion setter has greatly reduced ability to affect an unambiguous CS- (e.g., ¹⁶). In Experiment 2, one of three tests found congruent results that a 1st-order negative occasion setter has greatly reduced ability to affect responding to an unambiguous CS+ (e.g., ²²). These less consistent results for 1st-order negative occasion setters likely stem from the relative difficulty of learning negative vs positive 1st-order occasion setting, which has been shown elsewhere (^{7,13,23}) and in the present experiments.

Second, one of our critical novel hypotheses was that a 2nd-order occasion setter would not affect *unambiguous* lower-order stimuli (i.e., 1st-order occasion setters, CSs) but would affect *ambiguous* lower-order stimuli. This hypothesis was supported in both experiments for all analyses. Importantly, we tested the same 2nd-order occasion setter/1st-order occasion setter/CS stimulus arrangements – the only difference being the conditions of testing, which occurred while the 1st-order occasion setter/CS combination was *unambiguous* and later when *ambiguous*. This experimental manipulation allows us to causally determine that ambiguity increased the effects of the 2nd-order occasion setter on the lower-order 1st-order occasion setter/CS.

However, this does not mean that ambiguity is the only determinant of higher-order learning, and we do not claim that it is. Rather, we claim that lower-order stimulus ambiguity is a *necessary* but not *sufficient* condition for 2nd-order occasion setting to occur and that other specific factors help determine which ambiguous stimuli will be affected by an occasion setter. One example of these specific factors is lower-order stimulus training history (¹⁶). For example, a CS+ that undergoes acquisition unambiguously predicts the US, but if it later undergoes extinction, it becomes ambiguous. However, studies show mixed results in the ability of a 1st-order occasion setter to affect responding to a trained/extinguished CS+ (^{16,24,25}). In this example, during acquisition, the CS+ is unambiguously paired with the US, so its excitatory training does not involve a 1st-order positive occasion setter; after extinction, responding to the CS+ does not consistently increase when presented with a 1st-order positive occasion setter (despite the CS's ambiguity) likely because the CS+ was never trained with a 1st-order positive occasion setter. As another example, when OS1_A/CS_B and OS1_C/CS_D are trained (where “OS1” is a 1st-order occasion setter and the subscript letters are stimuli), the OS1 from one training will affect the CS from another training (e.g., OS1_A will affect CS_D), but this effect ranges from being partial (^{16,26}) to complete (²⁷). A pure ambiguity perspective would predict the 1st-order occasion setters would completely transfer their effects to a separately trained ambiguous CS. One explanation for the inconsistency in results could be stimulus generalization, where

within-OS1 similarity and within-CS stimulus similarity could facilitate transfer (²²) but dissimilarity could hinder it. Alternatively, the modulation account posits that 1st-order occasion setters affect the CS/US association itself (⁹), so when either the CS or US is changed, decrements in responding occur (^{28,29}). This collective evidence could be interpreted as failures of the ambiguity hypothesis. More likely, they indicate other factors built upon ambiguity (e.g., training history, stimulus similarity) influence when ambiguity has an effect. Importantly, there are ample studies demonstrating that 1st-order occasion setters affect ambiguous CSs (^{16,26,27}) but none that definitively show that they affect *unambiguous* CSs (the very few studies that do show this effect can be explained by less interesting phenomena, such as stimulus generalization or second-order conditioning; e.g., ³⁰). Thus, it seems that ambiguity is a *necessary* but not *sufficient* condition for 2nd order occasion setting to occur.

We are not the first to posit this ambiguity hypothesis (^{2,4-6,31}) or to experimentally assess it (²⁶). However, we are the first to extend it to 2nd-order occasion setting, investigate its robustness across learning hierarchies, and to create a theoretical and computational model operationalizing CS ambiguity and 1st-order occasion setter ambiguity. The data from the present experiments supports our computational model and its unique aspect of operationalizing CS ambiguity (i.e., CS ambiguity = direct excitation * direct inhibition) and 1st-order occasion setter ambiguity (i.e., OS1 ambiguity = positive OS1 * negative OS1) as the gateway to higher-order learning. Additionally, in terms of both WAIC and R², our 2nd-order occasion setting model outperformed our 1st-order occasion setting model, and both models greatly outperformed a direct associations model (i.e., Rescorla-Wagner). Thus, our model-free and model-based results converged, and both suggest that 2nd-order occasion setting, 1st-order occasion setting, and direct associations were learned.

It should be noted that both configural and modulation theories can account for the results of our experiments. Modulation is thought to be a more accurate account of the learning underlying occasion setting, which is why we conceptually and computationally focus on the modulation perspective (⁹). However, there is much overlap in predicted behavior between the two theories (⁹). A primary method to distinguish configuration vs modulation is manipulating the specificity of the CS/US association (e.g., using multiple CSs and USs; ⁹). Since this was not the goal of the present study, our design cannot distinguish between these two theories.

The present results provide numerous clinical implications across many disorders that are embedded with Pavlovian processes (e.g., anxiety, substance use). For example, individuals with anxiety disorders have deficits in working memory (³²⁻³⁴), and learning occasion setting is demanding of working memory – especially 2nd-order occasion setting. In addition to anxious individuals' deficits in discriminating safety from danger with direct associations (³⁵⁻³⁷) and 1st-order occasion setting (⁷), perhaps anxious individuals have deficits in learning 2nd-order occasion setting partly due to working memory deficits. This may be further compounded by anxious individuals' intolerance of uncertainty (^{38,39}), of which there is relatively more when learning higher-order vs lower-order Pavlovian associations. Thus, anxious individuals may show relatively greater fear of 1st- and 2nd-order occasion setting than non-anxious individuals. Second, our model claims that 1st-order occasion setters can be ambiguous (as has been shown elsewhere; e.g., ^{13,21}), and it argues that CS responding will be minimal if the CS has direct inhibition, is presented with 1st-order negative occasion setters, and is absent of 1st-order positive occasion setters. This is relevant for anxiety disorders, as conventional exposure therapy focuses primarily on direct CS extinction with some effort to increase 1st-order negative occasion setters via context variability (⁴⁰), but there is no explicit emphasis on 1st-order positive occasion setter extinction. Extinction of 1st-order positive occasion setting has been demonstrated elsewhere (^{19,20,41}), and, from our model's perspective, a 2nd-order negative occasion setter is the stimulus/context which determines how the extinguished 1st-order positive occasion setter will affect the CS/US association. This may suggest that exposure therapists should not only encourage CS extinction, but 1st-order positive occasion setter extinction, as well – the latter of which will be under the influence of a 2nd-order negative occasion setter.

One limitation of our experiments is that our sample was collected online, so an experimenter was not present to observe data collection. However, we included multiple data quality checks to ensure data validity (e.g., instructions quizzes), and others have shown validity of online data using the specific service

we used (i.e., Prolific; ⁴²). Importantly, the data shows that participants learned the complex discriminations appropriately, providing empirical evidence that our data is valid.

In conclusion, our experiments were the first to explicitly demonstrate the existence of 2nd-order occasion setting. We experimentally showed that lower-order stimulus ambiguity (i.e., conditional stimuli, 1st-order occasion setters) was necessary for higher-order learning (i.e., 1st- and 2nd-order occasion setting) to occur, and we argue that this is true in most if not all cases. Additionally, our computational model showed validity in predicting direct associations, 1st-order occasion setting, and 2nd-order occasion setting. These experiments are important for drawing research attention to an additional layer of Pavlovian learning (2nd-order occasion setting), as well as informing the treatment of anxiety disorders and other Pavlovian-based disorders (e.g., substance use).

Methods

Participants

Prolific⁽⁴²⁾ was used to recruit and collect human participant data online (Experiment 1: final N = 58; Experiment 2: final N = 67). Eligibility criteria included being age 18-65, healthy or corrected vision, United States residents, English fluent, no hearing difficulties, and a Prolific approval rating of $\geq 95\%$; participants were only allowed to participate in one of the experiments. Across both experiments, demographics information included gender (53.60% female, 45.60% male, 0.80% agender), age (mean = 30.18, SD = 10.88, min = 18, max = 63), and ethnicity (10.40% Black or African-American, 8% Central/East Asian (e.g., Chinese, Japanese, Korean), 4.80% Hispanic or Latin(x), 6.40% South Asian (e.g., Indian, Pakistani, Sri Lankan), 62.40% White, and 8% Multiracial). Participants were paid \$19.42 in Experiment 1 and \$19.06 in Experiment 2 for completing the study. This amount was achieved by US presentations at the end of reinforced trials, where each US was a \$0.12 USD increase in payment (as well as \$2.50 for completing questionnaires). This study was deemed exempt by the California Institute of Technology Institutional Review Board, and all participants provided informed consent prior to commencing the study.

Additionally, prior to data collection, we pre-registered each study (Experiment 1: <https://osf.io/n2c6v>, Experiment 2: <https://osf.io/hxcfs>). Based on our pre-registered exclusion criteria (i.e., automatic/invariant responding), we excluded one participant from each Experiment 1 and 2; also, an additional participant from Experiment 2 was excluded because of technical difficulties.

Design

There were no between-subjects conditions; within-subjects conditions included stimuli with direct associations with the US (i.e., CSs), 1st-order occasion setters, and 2nd-order occasion setters. Experiment 1 was a 2nd-order *negative* occasion setting design, which included 1st-order positive occasion setting and 2nd-order negative occasion setting. Experiment 2 was a 2nd-order *positive* occasion setting design, which included 1st-order negative occasion setting and 2nd-order positive occasion setting. Thus, Experiments 1 and 2 are mirror opposites of each other. Both experiments included CSs with direct associations with the US. Our dependent variable was US expectancy, measured at the end of every trial.

Materials and Apparatus

The Pavlovian conditioning procedure was programmed using PsychoPy 2020.1.3. All learning stimuli (CSs, 1st-order occasion setters, 2nd-order occasion setters) were 4sec audio or visual stimuli. When multiple stimuli were presented within the same trial, they were presented *serially* with a 4sec inter-stimulus interval (ISI) between them. Serial presentation (rather than simultaneous presentation) is critical in most cases to learn occasion setting rather than direct associations^(45,46). The US was a 1.25sec audio/visual stimulus showing an image of a gold coin with “12¢” written on it, confetti surrounding it, and an audio cash register sound (i.e., “cha-ching!”). Inter-trial intervals (ITIs) were 1.25sec. ITIs and ISIs included a fixation cross, which was also displayed uninterrupted during audio stimuli. All trials ended with a US expectancy rating, which had no time constraint (see Figure 2).

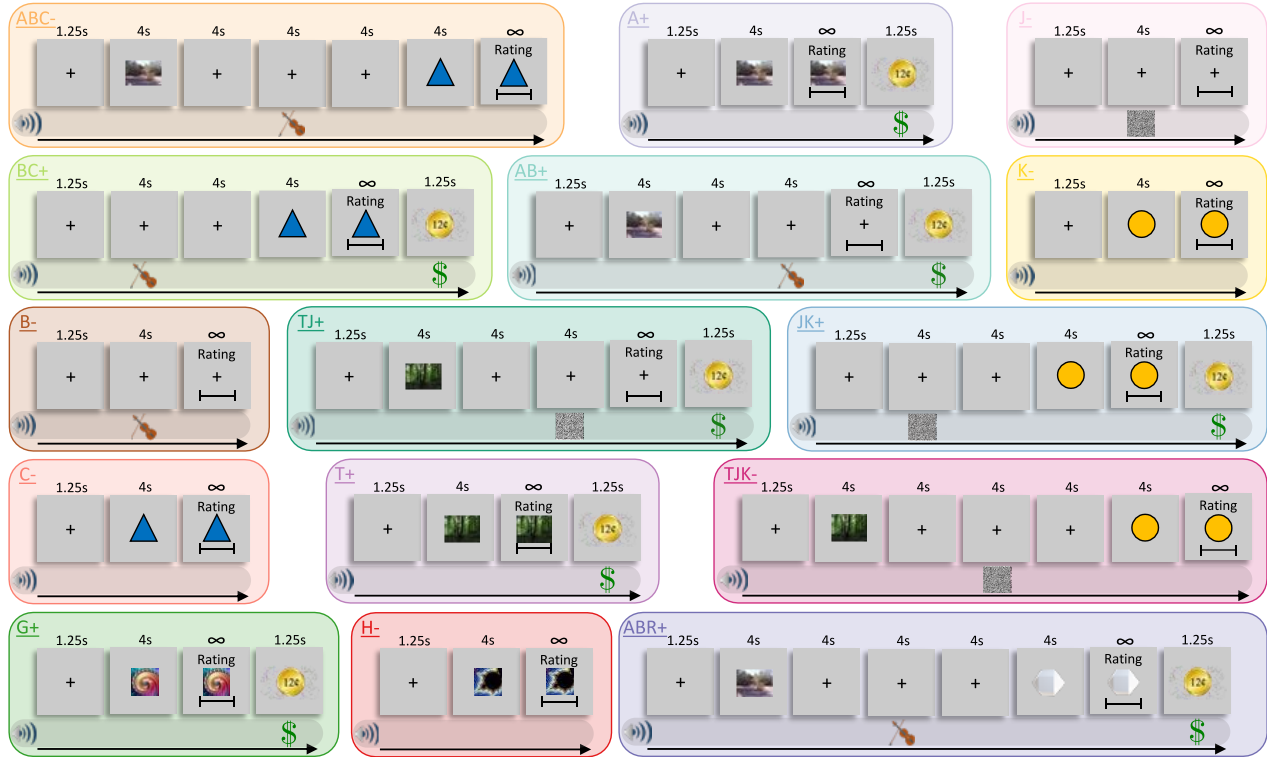


Figure 2. Experiment 1 (2nd-Order Negative Occasion Setting) Trial Design. Each colored box represents a trial type presented in the experiment. Gray boxes represent what was shown visually on screen. Inter-trial intervals (ITIs) and inter-stimulus intervals (ISIs) included a gray screen with a fixation cross (“+”). Duration of each trial component is shown at top of each trial type. Rating slide is shown in abbreviated form, and visual analog scale was used to rate US Expectancy. Auditory stimuli are indicated below slides in horizontal auditory band (). Violin symbol indicates violin sound (), static screen indicates white noise (), and dollar sign indicates cash register sound (). None of the auditory symbols were shown on screen during the experiment. Black arrow pointing to the right () for each trial type indicates chronological sequence during trials. All stimuli are counterbalanced across participants within stimulus category: G/H (Unambiguous CSs), C/K (Ambiguous CSs), B/J (1st-order occasion setters), and A/T (2nd-order occasion setters). All trial types for Experiment 1 are shown except TJR+, which is identical to ABR+, but T and J stimuli are substituted for A and B. Experiment 2 (2nd-order positive occasion setting) is a mirror image design of Experiment 1 in which all trial types reinforced in Experiment 1 were not reinforced in Experiment 2, and all trial types not reinforced in Experiment 1 were reinforced in Experiment 2. The only exceptions are G+ and H-, which remained a CS+ and CS-, respectively, in each study.

Unambiguous CSs (i.e., G+, H-) were images of fractals, ambiguous CSs (Experiment 1: C, K; Experiment 2: F, N) were images of a blue triangle and orange circle, OS1s (Experiment 1: B, J; Experiment 2: E, M) were a violin sound and white noise sound, and OS2s (Experiment 1: A, T; Experiment 2: D, U) were images of a desert and forest. Within each category, stimuli were counterbalanced across participants. Using different stimulus modalities (e.g., auditory, visual) between hierarchical levels facilitates distinction between direct and OS learning (¹¹). This is likely because different modalities help distinguish the occasion setter and CS in physical properties and in associative meaning. Given that we only had two modalities to use, we made the 2nd-order occasion setting level visual (to distinguish from 1st-order occasion setting) but categorically different from the CS images (i.e., context images vs shapes/fractals). An additional unambiguous CS (Experiment 1: R; Experiment 2: S) was an image of a three-dimensional white gem; this stimulus was used to facilitate 2nd-order occasion setting (see Supplemental Materials). Notably, the above list is the *primary* function of each stimulus, but a given stimulus may have had more than one hierarchical meaning. For instance, each “1st-order occasion setter” listed above was also a CS with a direct association with the US, and this was modulated by an “2nd-order occasion setter” listed above, which acted as a 1st-order occasion setter in that case (e.g., Experiment 1: “B” was a 1st-order occasion setter (C-, BC-) but also a CS (B-, AB+); “A” was a 2nd-order occasion setter (ABC-) but also a 1st-order occasion setter (B-, AB-) and a CS (A+)). This allowed us to test the specific hierarchical functions of each stimulus and determine whether independence between hierarchical levels is learnable. Lastly, the following is a list of congruent stimuli between each Experiment (listed as Experiment 1/Experiment 2): A/D, B/E, C/F, J/M, K/N, R/S, T/U. G+ and H- were identical across both studies.

US expectancy. Participants used a visual analog scale to rate, “How certain are you that you are about to receive a bonus payment?” The values ranged from 1 = “Certain No Bonus”, 3 = “Completely Uncertain,” and 5 = “Certain Yes Bonus.” The visual analog scale did not display numerical values, but it displayed the anchors mentioned above. US expectancy was measured at the end of every trial using the mouse to click on the scale with unlimited time to respond.

Procedure

Participants attended one experimental session online lasting approximately 1 hour 45 minutes, where they provided informed consent, completed questionnaires, and completed the Pavlovian learning experiment. In the experiment (see Table 1), participants were informed, “Your goal in this experiment is to learn which sounds and images predict receiving bonus payments.” During Training and Reminder phases, participants experienced US (non)reinforcement, which resulted in learning and real increases in their payment. Importantly, to maintain the Pavlovian nature of the experiment (rather than instrumental), participant responses did not affect their payment. During Transfer Test, participants were not informed whether they received the US, which was accomplished by using an image of a curtain to cover the location on screen where the US image would otherwise occur and muting the US sound. This curtain/muted modification was done so that no learning and no reinforcement/non-reinforcement would occur during Transfer Test, allowing us to test the underlying learning processes that occurred during training with many Transfer Test trial types but without learning via the non(reinforcement) of the critical transfer test trials. In sum, Training was conducted for participants to learn *which* stimuli predict (non)reinforcement, and Transfer Test was conducted to investigate *how* participants learned which stimuli predict (non)reinforcement.

Table 1a. Experiment 1 (2nd-Order NOS) Training Sequence

1st Training				
<i>Unambiguous</i> POS1 (JK)	<i>Ambiguous</i> POS1 (BC)	<i>NOS2</i> (ABC)	<i>Direct</i> (ABR)	<i>Direct</i> (G, H)
10 each: J-, K-, JK+	5 each: B-, C-, BC+	10 ABC-, 5 each: B-, C-, BC+, A+, AB+	10 ABR+	10 each: G+, H-
1st Reminder		1st Transfer Test		
10 each: ABC-, J-, K-, JK+, G+, H-, ABR+. 5 each: B-, C-, BC+, A+, AB+	<i>NOS2 to Unambiguous</i> POS1 & CS+	<i>Ambiguous POS1 to</i> CS-	<i>Unambiguous</i> POS1 to CS-	<i>NOS2/POS1 to</i> CS+ & CS-
	3 each: ABC1, ABR, AJK1, ABG	3 each: BC, H, BH	3 each: JK1, H, JH	3 each: G, H, AG, AH
2nd Training		2nd Reminder	2nd Transfer Test	
<i>NOS2</i> (TJK)	<i>Direct</i> (TJR)		<i>NOS2 to Ambiguous POS1</i>	
10 TJK-, 5 each: J-, K-, JK+, T+, TJ+	10 TJR+	9 each: TJK-, T+. 3 each: ABC-, BC+, C-, A+, JK+, K-	3 each: ABC2, TJK2, JK2, AJK2	

Table 1b. Experiment 2 (2nd-Order POS) Training Sequence

1st Training				
<i>Unambiguous</i> NOS1 (MN)	<i>Ambiguous</i> POS1 (EF)	<i>NOS2</i> (DEF)	<i>Direct</i> (DES)	<i>Direct</i> (G, H)
10 each: M+, N+, MN-	5 each: E+, F+, EF-	10 DEF+. 5 each: E+, F+, EF-, D-, DE-	10 DES-	10 each: G+, H-
1st Reminder		1st Transfer Test		
10 each: DEF+, M+, N+, MN-, G+, H-, DES-. 5 each: E+, F+, EF-, D-, DE-	<i>POS2 to Unambiguous</i> NOS1 & CS-	<i>Ambiguous NOS1</i> to CS+	<i>Unambiguous</i> NOS1 to CS+	<i>POS2/NOS1 to CS-</i> & CS+
	3 each: DEF1, DES, DMN1, DEH	3 each: EF, G, EG	3 each: MN1, G, MG	3 each: G, H, DG, DH
2nd Training		2nd Reminder	2nd Transfer Test	
<i>POS2</i> (UMN)	<i>Direct</i> (UMS)		<i>POS2 to Ambiguous NOS1</i>	
10 UMN+. 5 each: M+, N+, MN-, U-, UM-	10 UMS-	9 each: UMN+, U-. 3 each: DEF+, EF-, F+, D-, MN-, N+	3 each: DEF2, UMN2, MN2, DMN2	

Table 1. Experiment Training Sequence. Experiments 1 and 2 are mirror images of each other in which reinforced and non-reinforced trials are swapped. Within sub-tables **a** and **b**, chronological training sequence occurs from top-left to bottom-right. Major phases include 1st Training, 1st Reminder, 1st Transfer Test, 2nd Training, 2nd Reminder, and 2nd Transfer Test. "NOS" and "POS" refer to negative and positive occasion setting, respectively. The association numbers (e.g., NOS1, NOS2) refer to 1st-order and 2nd-order occasion setting, respectively. Trial numbers for each stimulus are indicated; Transfer Tests include 3 trials of each stimulus per phase. Within each phase, trial sequence is randomized in minimal divisible blocks (e.g., 1st Reminder: 2 trials each for ABC-, J-, etc. and 1 trial each of B-, C-, etc.). Within 1st Transfer Test, each test predicts *lack* of transfer to lower-order *unambiguous* stimuli (i.e., 2nd OS will not transfer to unambiguous OS1 or unambiguous CS). Within Transfer Test 2, we predict *presence* of transfer within 2nd OS hierarchical level to *ambiguous* OS1, which will be evaluated by comparing the same stimuli when OS1 was *unambiguous* (Transfer Test 1) and *ambiguous* (Transfer Test 2) (i.e., Experiment 1: AJK2 vs AJK1; Experiment 2: DMN2 vs DMN1).

Data Analysis

We used Stata 15.1 multilevel modeling for inferential statistics. For US expectancy and SCR during the Training phase, Level 1 predictors were Stimulus, Linear Slope (e.g., Trial Blocks 1-5), and Quadratic Slope (e.g., Trials Blocks 1-5). If the Quadratic Slope was not significant, it was removed from the model and re-run as a linear model. If the Linear Slope was not significant, it was removed and collapsed across Stimulus. For Transfer Test, the Level 1 predictor was Stimulus using the average of all three trials from a given block. For Transfer Test, we calculated difference scores between the relevant stimuli and conducted t-tests with those difference scores.

Furthermore, computational modeling was conducted using Python 3.7.6. We used computational modeling to evaluate our theoretical model's fit with the data. In our models, the only free parameter was learning rate (i.e., α). In short, our model separately calculates direct excitation (i.e., "direct" meaning associations between a CS and US), direct inhibition, 1st-order positive occasion setting (i.e., where the OS1 modulates the current CS/US association), 1st-order negative occasion setting, 2nd-order positive occasion setting (i.e., where the OS2 modulates the effect of OS1 on the CS/US association), and 2nd-order negative occasion setting. It then predicts behavioral responding by adding together direct excitation, 1st-order positive occasion setting, and 2nd-order positive occasion setting, and subtracting direct inhibition, 1st-order negative occasion setting, and 2nd-order negative occasion setting into the variable R (i.e., "responding"). Like the Rescorla-Wagner model⁽¹⁰⁾, the normal bounds of learning/responding are -1 to 1, where $R > 0$ is excitation and $R < 0$ is inhibition. The Supplementary Materials provide details about the model. Additionally, we have included an html file in our Supplementary Materials (titled "Zbozinek et al 2nd-Order Occasion Setting Formulas.html") that provides an interactive figure for readers to engage with. With this figure, readers can manipulate the variables of the model using sliding scales and view the output. We highly suggest the readers use this file to better understand the model and formulas.

In our models, learning rate (α) is a free parameter. We ran one model at each hierarchical level: Rescorla-Wagner (direct associations only), our 1st-order occasion setting model (direct associations and 1st-order occasion setting), and our 2nd-order occasion setting model (direct associations and 1st- and 2nd-order occasion setting). The Rescorla-Wagner model had a general α (one α for all stimuli). Within our occasion setting models, we ran two different α models: General (one α for all stimuli) and Stimulus-Specific (one α for each unique trained stimulus association; 15 total). For general model fitting, we allowed γ_1 (i.e., CS ambiguity), γ_2 (i.e., OS1 ambiguity), and λ (i.e., absence of an expected US) to scale according to their formulas (see Supplemental Materials). The natural scaling of these variables slows down predicted learning since they are used in the trial-by-trial Δ formulas to predict change in learning. Thus, they inflate α values for the relevant stimuli (i.e., 1st- and 2nd-order occasion setters and inhibitory stimuli). For our primary analyses, we calculated γ_1 , γ_2 , and λ according to their formulas. Secondly, to obtain the most accurate α values, we ran separate models in which we fixed γ_1 , γ_2 , and λ to 1, allowing all individual differences in learning rate across stimuli and participants to be funneled into α .

All models were fit using a hierarchical Bayesian approach, assuming subject-level parameters were drawn from group-level distributions, with parameters estimated using variational inference implemented in PyMC3 with 40,000 iterations. To evaluate α parameter recovery, we separately simulated random α values ranging 0-1 and recovered them using our model with our sample size. In both, the correlation of simulated vs recovered α parameters in our best-fitting model was very high ($r_s > .998$; see Supplementary Materials). Model comparison was performed using Watanabe-Akaike Information Criterion (WAIC) scores⁽⁴⁷⁾, which provides a goodness of fit measure for Bayesian models, penalizing for increasing numbers of free parameters in the model (lower scores indicate better model fit). We additionally estimated R^2 values for training trials (which encompasses both Training and Reminder phases) using scikit, where the best score is 1, and negative scores indicate the model fitting arbitrarily worse than a constant model. Ultimately, we evaluated models based on WAIC scores, R^2 values, and robustness across both experiments.

References

1. Zbozinek, T. D. & Craske, M. G. Pavlovian extinction of fear with the original conditional stimulus, a generalization stimulus, or multiple generalization stimuli. *Behav. Res. Ther.* **107**, 64–75 (2018).
2. Fraser, K. M. & Holland, P. C. Occasion setting. *Behav. Neurosci.* **133**, 145–175 (2019).
3. Trask, S., Thrailkill, E. A. & Bouton, M. E. Occasion setting, inhibition, and the contextual control of extinction in Pavlovian and instrumental (operant) learning. *Behav. Processes* **137**, 64–72 (2017).
4. Bouton, M. E. Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychol. Bull.* **114**, 80–99 (1993).
5. Bouton, M. E. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol. Psychiatry* **52**, 976–986 (2002).
6. Rosas, J. M., Aguilera, J. E. C., Álvarez, M. M. R. & Abad, M. J. F. Revision of Retrieval Theory of Forgetting: What does Make Information Context-Specific? *Int. J. Psychol. Psychol. Ther.* **6**, 147–166 (2006).
7. Zbozinek, T. D., Wise, Toby, Qi, Song, Fanselow, M & Mobbs, D. Occasion Setting, Anxiety, and their Computation in Human Fear Conditioning. (2020)
doi:<https://doi.org/10.31234/osf.io/7xse6>.
8. Swartzentruber, D. Perspectives on modulation: Modulator- and target-focused views. in *Occasion setting: Associative learning and cognition in animals* 167–197 (American Psychological Association, 1998). doi:10.1037/10298-006.
9. Bonardi, C., Robinson, J. & Jennings, D. Can existing associative principles explain occasion setting? Some old ideas and some new data. *Behav. Processes* **137**, 5–18 (2017).

10. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Class. Cond. II Curr. Res. Theory* **2**, 64–99 (1972).
11. Holland, P. C. Acquisition and transfer of conditional discrimination performance. *J. Exp. Psychol. Anim. Behav. Process.* **15**, 154–165 (1989).
12. Holland, P. C. Transfer of negative occasion setting and conditioned inhibition across conditioned and unconditioned stimuli. *J. Exp. Psychol. Anim. Behav. Process.* **15**, 311–328 (1989).
13. Holland, P. C. Transfer of control in ambiguous discriminations. *J. Exp. Psychol. Anim. Behav. Process.* **17**, 231–248 (1991).
14. Baeyens, F., Vansteenwegen, D., Hermans, D., Vervliet, B. & Eelen, P. Sequential and simultaneous feature positive discriminations: Occasion setting and configural learning in human Pavlovian conditioning. *J. Exp. Psychol. Anim. Behav. Process.* **27**, 279–295 (2001).
15. Baeyens, F. *et al.* Simultaneous and sequential Feature Negative discriminations: Elemental learning and occasion setting in human Pavlovian conditioning. *Learn. Motiv.* **35**, 136–166 (2004).
16. Holland, P. C. Transfer after serial feature positive discrimination training. *Learn. Motiv.* **17**, 243–268 (1986).
17. Holland, P. C. Acquisition and transfer of occasion setting in operant feature positive and feature negative discriminations. *Learn. Motiv.* **22**, 366–387 (1991).
18. Arnold, H. M., Grahame, N. J. & Miller, R. R. Higher order occasion setting. *Anim. Learn. Behav.* **19**, 58–64 (1991).

19. Franssen, M. *et al.* Reinstatement after human feature-positive discrimination learning. *Behav. Processes* **137**, 73–83 (2017).
20. Rescorla, R. A. Extinction of facilitation. *J. Exp. Psychol. Anim. Behav. Process.* **12**, 16 (1986).
21. Holland, P. C. & Reeve, C. E. Acquisition and transfer of control by an ambiguous cue. *Anim. Learn. Behav.* **19**, 113–124 (1991).
22. Holland, P. C. & Lamarre, J. Transfer of inhibition after serial and simultaneous feature negative discrimination training. *Learn. Motiv.* **15**, 219–243 (1984).
23. Holland, P. C., Lamoureux, J. A., Han, J.-S. & Gallagher, M. Hippocampal lesions interfere with Pavlovian negative occasion setting. *Hippocampus* **9**, 143–157 (1999).
24. Swartzentruber, D. & Rescorla, R. A. Modulation of trained and extinguished stimuli by facilitators and inhibitors. *Anim. Learn. Behav.* **22**, 309–316 (1994).
25. Morell, J. R. & Davidson, T. L. Transfer across unconditioned stimuli in serial feature discrimination training. *J. Exp. Psychol. Anim. Behav. Process.* **28**, 83–96 (2002).
26. de Brugada, I., Garcia-Hoz, V., Bonardi, C. & Hall, G. Role of stimulus ambiguity in conditional learning. *J. Exp. Psychol. Anim. Behav. Process.* **21**, 275–284 (1995).
27. Davidson, T. L. & Rescorla, R. A. Transfer of facilitation in the rat. *Anim. Learn. Behav.* **14**, 380–386 (1986).
28. Bonardi, C., Bartle, C. & Jennings, D. US specificity of occasion setting: Hierarchical or configural learning? *Behav. Processes* **90**, 311–322 (2012).
29. Bonardi, C. & Ward-Robinson, J. Occasion Setters: Specificity to the US and the CS–US Association. *Learn. Motiv.* **32**, 349–366 (2001).

30. Bonardi, C. Occasion setting without feature-positive discrimination training. *Learn. Motiv.* **23**, 343–367 (1992).
31. Swartzentruber, D. Modulatory mechanisms in Pavlovian conditioning. *Anim. Learn. Behav.* **23**, 123–143 (1995).
32. Darke, S. Anxiety and working memory capacity. *Cogn. Emot.* **2**, 145–154 (1988).
33. Shackman, A. J. *et al.* Anxiety selectively disrupts visuospatial working memory. *Emotion* **6**, 40–61 (2006).
34. Eysenck, M., Payne, S. & Derakshan, N. Trait anxiety, visuospatial processing, and working memory. *Cogn. Emot.* **19**, 1214–1228 (2005).
35. Duits, P. *et al.* Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depress. Anxiety* **32**, 239–253 (2015).
36. Lissek, S. *et al.* Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behav. Res. Ther.* **43**, 1391–1424 (2005).
37. Craske, M. G. *et al.* Elevated responding to safe conditions as a specific risk factor for anxiety versus depressive disorders: Evidence from a longitudinal investigation. *J. Abnorm. Psychol.* **121**, 315–324 (2012).
38. Carleton, R. N. *et al.* Increasingly certain about uncertainty: Intolerance of uncertainty across anxiety and depression. *J. Anxiety Disord.* **26**, 468–479 (2012).
39. Jensen, D., Cohen, J. N., Mennin, D. S., Fresco, D. M. & Heimberg, R. G. Clarifying the unique associations among intolerance of uncertainty, anxiety, and depression. *Cogn. Behav. Ther.* **45**, 431–444 (2016).
40. Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T. & Vervliet, B. Maximizing exposure therapy: An inhibitory learning approach. *Behav. Res. Ther.* **58**, 10–23 (2014).

41. Miller, R. R. & Oberling, P. Analogies between occasion setting and Pavlovian conditioning. in *Occasion setting: Associative learning and cognition in animals* 3–35 (American Psychological Association, 1998). doi:10.1037/10298-001.
42. Peer, E., Brandimarte, L., Samat, S. & Acquisti, A. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *J. Exp. Soc. Psychol.* **70**, 153–163 (2017).
43. Pérez, O. D., San Martín, R. & Soto, F. A. Exploring the effect of stimulus similarity on the summation effect in causal learning. *Exp. Psychol.* **65**, 183–200 (2018).
44. Melchers, K. G., Shanks, D. R. & Lachnit, H. Stimulus coding in human associative learning: Flexible representations of parts and wholes. *Behav. Processes* **77**, 413–427 (2008).
45. Holland, P. C. Differential effects of reinforcement of an inhibitory feature after serial and simultaneous feature negative discrimination training. *J. Exp. Psychol. Anim. Behav. Process.* **10**, 461–475 (1984).
46. Holland, P. C. Temporal determinants of occasion setting in feature-positive discriminations. *Anim. Learn. Behav.* **14**, 111–120 (1986).
47. Watanabe, S. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *J. Mach. Learn. Res.* **11**, 3571–3594 (2010).

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1911441 granted to Tomislav Zbozinek, PhD under the supervision of Dean Mobbs, PhD and Michael Fanselow, PhD. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Author Contributions

Study Conceptualization: Tomislav Zbozinek

Experiment Design: Tomislav Zbozinek, Omar Perez, Dean Mobbs, Michael Fanselow

Funding Acquisition: Tomislav Zbozinek, Dean Mobbs, Michael Fanselow

Supervision: Dean Mobbs, Michael Fanselow

Experiment Programming: Tomislav Zbozinek

Data Acquisition: Tomislav Zbozinek

Data Analysis: Tomislav Zbozinek

Data Visualization: Tomislav Zbozinek

Occasion Setting Formula Development: Tomislav Zbozinek

Occasion Setting Formula Review and Feedback: Omar Perez, Michael Fanselow, Toby Wise

Computational Modeling: Tomislav Zbozinek, Toby Wise

Writing – Original Draft: Tomislav Zbozinek

Writing – Review and Editing: all authors